A Photo-Sketch and Sketch-Photo Synthesis using Markov Random Fields

Patrick Buhagiar

Supervisor: Dr Ing. Reuben A. Farrugia

Faculty of ICT

University of Malta

May 2015

Submitted in partial fulfillment of the requirements for the degree of B.Sc. (Hons.) Computing Science

Contents

List of Figures

List of Tables

List of Abbreviations

AAM	Active Appearance Model
MRF	Markov Random Field
MAP	Maximum *a posteriori*
PCA	Principal Component Analysis
HMM	Hidden Markov Model
E-HMM	Embedded Hidden Markov Model
LLE	Local Linear Embedding
SIFT	Scale-invariant Feature Transform
LFDA	Local Feature-based Discriminant Analysis
HAOG	Histogram of Averaged Oriented Gradients
AOMs	Active Orientation Models
SSIM	Structural Similarity Index Metric
CUHK	Chinese University of Hong Kong
PSNR	Peak Signal-to-Noise Ratio
PHD	Pretty Helpful Development (Toolbox)

A Photo-Sketch and Sketch-Photo Synthesis using Markov Random Fields

Patrick Buhagiar*

Supervised by: Dr Ing. Reuben A. Farrugia

May 2015

Abstract: In this dissertation, a photo/sketch synthesis algorithm based on Markov random fields was explored with the aim of preparing sketches for automatic face recognition. Such an algorithm has useful applications in law enforcement where on several occasions police officers must rely on a sketch drawn from witness recollection. This dissertation adopts a set of training sketch/photo pairs to synthesize unseen sketches. To synthesize sketch/photo images, all images within this training set and the input image are transformed into a certain face template and divided into overlapping patches. The best matching candidate patches are chosen and their image pairs are modelled into a Markov network where the maximum *a posteriori* estimate is calculated. Several tests have shown that recognition performance improved when converting sketches and photos into the same modality and that overall sketch synthesis achieved the best result of 94.68% (rank one) recognition rate.

1 Introduction

Although face sketching is a simple tool for expressing face portraits, it has many useful applications in several areas such as law enforcement and digital entertainment. Sketches manage to capture the most important perceptual information with a number of strokes [UJdVL96]. Researchers have been using different types of face drawings such as line drawings [WT09] or pencil sketches in order to study how capable the human visual system is at face recognition. It was found that humans are even able to recognize persons from caricatures [BP91] and cartoon animated images [BHD+92].

This research eventually led to several proposals for computer based sketch synthesis systems. In [KTFM99], [ITO99], without the need for any learning algorithm, face features were extracted from photos and exaggerated slightly by some parameter in order to produce a realistic sketch. This system, along with those in [FTP99], [CXS+01] make use of face

*Submitted in partial fulfillment of the requirements for the degree of B.Sc. Computing Science (Hons.).

alignment algorithms, such as active appearance model (AAM) [CET01]. However they were limited to just line drawings. Compared to pencil sketches, line drawings are less expressive due to the absence of shading texture.

A common problem faced by law enforcement agencies is that in many cases, police officers must rely on a sketch based on witness recollection since an image of the suspect would not always be available. This sketch can be made in many forms such as a pencil sketch by a forensic artist or a sketch produced using composite software such as EFIT-V. Eventually, after creation, this sketch is released to the public in the hope that some individual would be able to identify the suspect. This process can be quite tedious and time consuming. It is worth emphasizing that the quality of the resulting sketch (in terms of resemblance) highly depends on the accuracy of the witness' description. Another instance where forensic sketches are used is when police officers must rely on poor quality surveillance images as illustrated in Figure 1.

Figure 1: An example of sketches drawn from poor surveillance images [Lov11].

A more efficient solution to the above problem is to use automated face recognition systems which are able to automatically recognize the person in a sketch by referring to a database of images such as driving license or ID card photos, thereby narrowing down the list of potential suspects. Such a system would not only save time, but also minimize the amount of subjective assessment. The need for effective and automatic recognition of sketches by reference to a photo database has attracted many researchers. Although photos and sketches might have similar structure, the texture will vary across the two modalities. Human skin texture is simplified when drawing sketches compared to that is captured by a camera.

Despite being a simplified version, in most cases we can often recognize a person from a

sketch. However, there is limited research work on face sketch recognition. This is mainly due to the fact that there was no large face sketch database available for experimental study, thus making it more difficult than photo based face recognition [WT09]. As a result, existing face recognition algorithms are designed to match photo based faces, not sketches [XGTL09].

Several methods for face sketch recognition have been introduced in the past few years. These are categorised into two general approaches [GS12]: *intra-modality* and *inter-modality*. The intra-modality approach consists of synthesizing a pseudo-photo from an input sketch (or vice versa) in order to perform automatic face recognition in the same modality. An inevitable step in this approach is photo/sketch synthesis. Alternatively in inter-modality approaches, face recognition is performed by extracting discriminative features that are invariant to photo and sketch modalities such as colour and texture descriptors.

In this dissertation, an intra-modality synthesis algorithm based on Markov random fields was explored in order to synthesize sketches for automatic face recognition systems. The performance of intra-modality approaches highly depends on the quality of the training set and the accuracy of the photo/sketch synthesis algorithms. What makes the problem non-trivial however is the fact that it is very difficult to define sketch/photo synthesis by simple rules or grammar [WT09]. Another reason is that sketches depend on eyewitness recollection and not on the data set provided. The synthesis algorithm is capable of converting between any two modalities including pencil sketches, computer generated composite sketches, line drawings, camera photos and even low/high resolution images (super-resolution). However, throughout this dissertation the main focus was restricted to sketch-to-photo and photo-to-sketch synthesis of face portraits for automated face recognition in the same modality.

2 Background

2.1 Bayesian Inference

A certain branch of mathematical probability deals with calculating the uncertainty of certain outcomes by combining general knowledge and observational evidence. This branch is called Bayesian probability theory and it is a very important technique for statistical purposes.

In real life situations one can notice that certain variables are dependent on other

variables. This is usually referred to as evidence. Everyday life presents many situations in which the gathering of evidence leads to a certain conclusion. These dependent variables can be modelled into a graphical structure called a Bayesian network. Formally, a Bayesian network, also known as a belief network, is a directed model of conditional dependence across a set of random variables [Bay]. It is represented as an acyclic directed graph such that nodes are variables and the edges show conditional dependencies. The lack of edges show conditional independence where there is no possibility of that variable depending on another. Building a belief network requires the inclusion of all important variables from the model and prior knowledge on how connections should be made, hence which variables depend on which.

Bayesian inference is the process of updating outcome probabilities based on the relationships and currently known evidence. In general, the joint probability distribution must first be calculated in order to correctly perform inference. When using a Bayesian network, evidence and observations are updated with regards to recent events. A few important notations are *prior* and *posterior* probabilities. Existing beliefs (or probabilities) within the model are called *prior* probabilities while beliefs computed after evidence and observations are called *posterior* probabilities.

2.2 Markov Random Field Modelling

Another branch of mathematical probability is Markov Random Fields (MRF) modeling where a set of random variables having a Markov property are formed into an undirected graph. The Markov property refers to the memory-less property of a stochastic process, such that the probability distribution of future states does not depend on the sequence of events that come before, but rather on just the current state.

MRF is used to establish probabilistic distributions of interacting labels. Several image processing and analysis problems can be described as labelling problems [Li09]. A labelling problem consists of a set of *sites* and a set of *labels*. A *site* represents a region in Euclidean space such as a pixel, corner, line segment or image patch. A set of sites in a 2-D image of size $n \times n$ for example is denoted as $S = \{(i, j)|1 \leq i, j \leq n\}$ where the indices i and j correspond to a pixel location on the image. The order of sites do not matter, however a certain relationship is kept between several sites through a *neighbourhood system*. Sites which do not present spatial regularity are called *irregular*, otherwise if they can be presented as a lattice, they are called *regular*.

A label describes the current state or a certain event happening on a particular site.

Labels can be either discrete or continuous. Continuous labels can take the shape of a vector or matrix value for example $L_c = \mathbb{R}^{a \times b}$ where a and b are dimensions. On the other hand, discrete labels consist of discrete values for example, in edge detection $L = \{edge, nonedge\}$ describes the set of possible labels. In essence a labelling of sites can be considered as a process for mapping the set of sites S to the set of labels L [Li09].

There are four categories to classify labelling problems. These categories are a combination of continuous or discrete labels and regular or irregular sites. Face sketch/photo synthesis falls under the category of regular sites with continuous labels, since sites refer to image patches while labels are from a real interval.

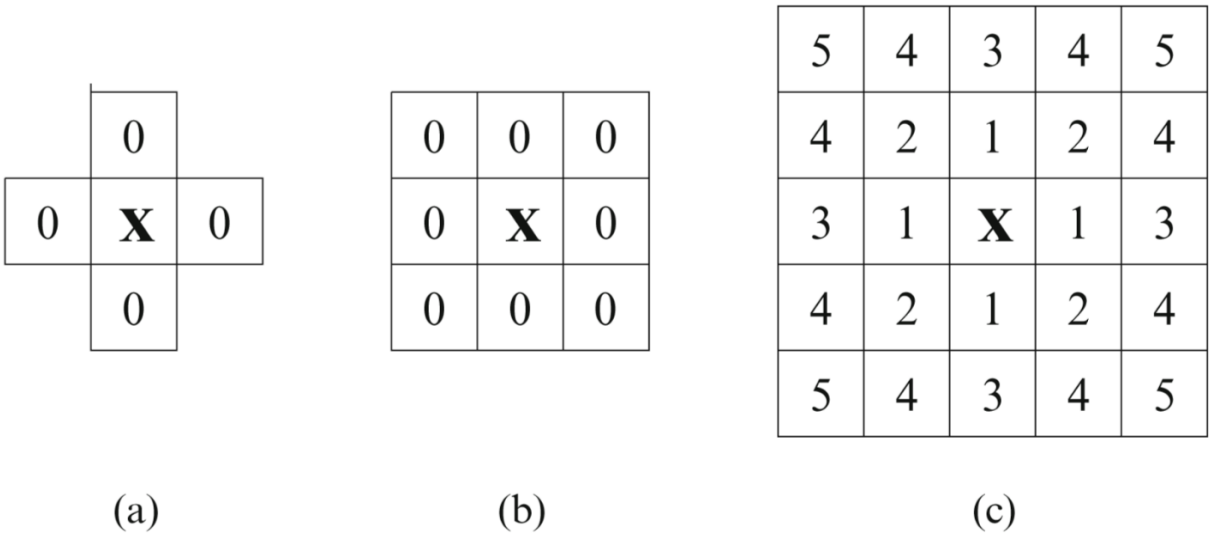

Figure 2: Neighbourhood System. Image from [Li09].

As mentioned earlier, *sites* are related via a neighbourhood system. For example, for a 2-D $n \times n$ image, each pixel will have a set of neighbouring pixels. The neighbourhood of a pixel varies according to its pixel location and its neighbourhood depth. Pixel location must be taken into consideration since edge pixels will have less neighbouring pixels than non-edge pixels. On the other hand, neighbourhood depth is defined as the set of neighbouring pixels within a certain radius of the concerning pixel. Rather than describing neighbourhood systems according to radius, a certain order is utilised. In a first order neighbourhood system, each pixel consists of 4 neighbours as in Figure 2(a). **X** denotes the concerned pixel while the zeros refers to its neighbours. Figure 2(b) shows a second order neighbourhood system, also known as the 8-neighbourhood system. The numbers from 1 to 5 in Figure 2(c) point to the outermost neighbouring pixels in the nth-order neighbourhood

system. The relationship between neighbouring pixels usually is represented by Euclidean distance however it might vary depending on the problem.

Markov random field theory is often used in conjunction with statistical decision methodologies. The most commonly used method to derive an estimate is finding the maximum *a posteriori* (MAP) probability. Similar to Bayesian probability theory, given some prior information, one can estimate parameters of a physical process. This prior information can come from previous empirical evidence. As an example, if we would like to estimate the value of parameter θ , the associated probabilities $P(\theta)$ are called prior probabilities [Awa07]. Bayes theorem shows how prior information can be incorporated to derive the posterior probability given by:

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} \tag{1}$$

where $P(\theta|x)$ is the likelihood term and $P(x)$ is a normalization term. With Bayesian inference, it is possible to find the optimal parameters which maximize the posterior probability. When prior information is available about θ, it is included in the prior distribution of θ. The maximum *a posteriori* estimate through Bayesian inference is defined as:

$$\arg\max\ P(\theta|x) = \arg\max\ P(x|\theta)P(\theta) \tag{2}$$

2.3 Markov Network for Sketch-Photo/Photo-Sketch Synthesis

It is necessary to estimate an underlying scene from a given image data. Calculating the maximum *a posteriori* (MAP) probability helps achieve an optimal solution. However in general it is very difficult to compute this probability without any approximations. A good solution would be to divide both the image and underlying scene into patches and assign each patch a node from the Markov network. Let x and y denote the estimate patch and the input patch respectively. The edges connecting the nodes are weighted according to their statistical dependency. Each scene x is connected to its corresponding image patch y and its neighbours as in Figure 3.

Similar to Bayes theorem, here the posterior probability is denoted as $P(x|y) = cP(x, y)$ where $c = \frac{1}{P(y)}$ is a constant over x. The best estimate $\hat{x}$ is the mode, or in other words the MAP probability, of the posterior probability $P(x|y)$.

There are two phases when solving a Markov network. The *learning phase* consists of learning network connection parameters from the training data. The *inference phase*

consists of estimating a particular unseen scene from some input image.

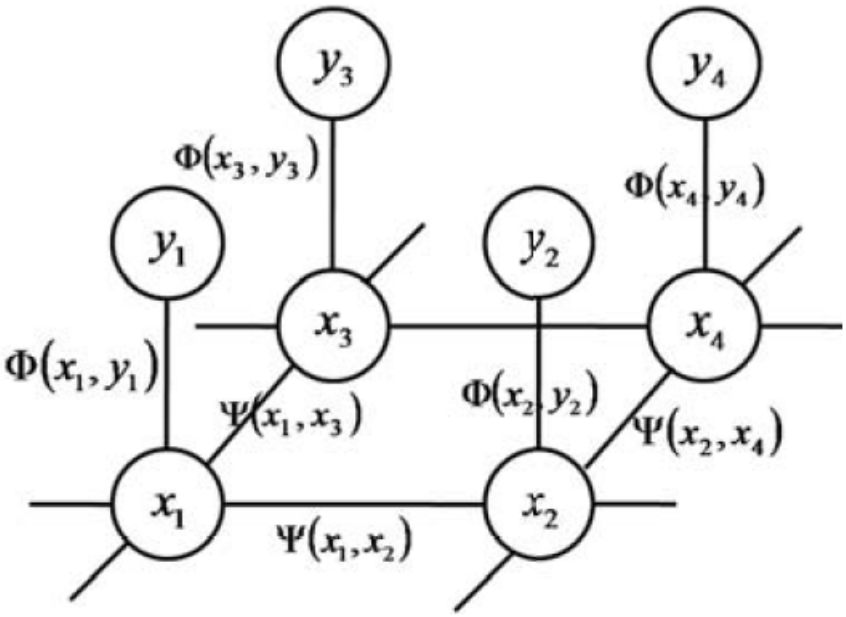

Figure 3: Graphical model of a Markov Network for Vision Problems. Image from [WT09].

The joint probability over estimate patch x and input patch y in Markov random fields (MRF) is written as:

$$P(x_1, x_2, ..., x_N, y_1, y_2, ...y_N) = \prod_{(i,j)} \psi(x_i, x_j) \prod_k \phi(x_k, y_k) \tag{3}$$

where ψ and ϕ are compatibility functions, (i, j) indicate neighbouring nodes i and j and N the number of nodes. The exact notation for ψ and ϕ is smoothness constraint and data constraint respectively. The data constraint term measures the fidelity between the input patch and target output while the smoothness constraints measures the local neighbourhood relationship of the target output. The estimate x is calculated by taking the maximum *a posteriori* (MAP):

$$\hat{x}_{j\,MAP} = \arg\max_{x_j} \arg \max_{[all\, x_i, i \neq j]} P(x_1, x_2, ..., x_N, y_1, y_2, ...y_N) \tag{4}$$

As an example, if we assume that the neighbours of node x_2 are x_1, x_4, x_7 and x_{10} in Figure

3, then the $\hat{x}_{2\,MAP}$ at node x_2 gives:

$$\hat{x}_{2\,MAP} = \arg \max_{x_1} \max_{x_2} \max_{x_4} \max_{x_7} \max_{x_{10}} P(x_1, x_2, ..., x_N, y_1, y_2, ...y_N) \tag{5}$$

$$\hat{x}_{2\,MAP} = \arg \max_{x_1} \max_{x_2} \max_{x_4} \max_{x_7} \max_{x_{10}} \phi(x_1, y_1)\phi(x_2, y_2)\phi(x_4, y_4)\phi(x_7, y_7)\phi(x_{10}, y_{10}) \psi(x_1, x_2)\psi(x_2, x_4)\psi(x_2, x_7)\psi(x_2, x_{10}) \tag{6}$$

$$\hat{x}_{2\,MAP} = \arg \max_{x_1} \phi(x_1, y_1) \max_{x_2} \phi(x_2, y_2)\psi(x_1, x_2) \max_{x_4} \phi(x_4, y_4)\psi(x_2, x_4) \max_{x_7} \phi(x_7, y_7)\psi(x_2, x_7) \max_{x_{10}} \phi(x_{10}, y_{10})\psi(x_2, x_{10}) \tag{7}$$

Equations 5, 6 and 7 break down the MAP estimation equation such that eventually each line of equation 7 is a local computation involving one node and its compatibility with the concerned node. In most situations, estimate patches x in the Markov model would consist of several candidate patches. In this case, the local computations in equation 7 takes into consideration the candidate patches that achieve the maximum possible probabilities. The choice of the ideal candidate patch will depend on how similar it is to the input patch and also how well it fits in with neighbouring patches. Since candidate patches are involved, the $x_{j\,MAP}$ probability must be calculated for every combination of patch candidate neighbourhoods. The candidate patch with highest joint probability combination is chosen as the estimate patch.

3 Literature Review

In the past few years intra-modality and inter-modality approaches have been introduced in order to solve the face sketch recognition problem. The intra-modality approach consists of converting images into the same modality through face sketch/photo synthesis. Alternatively in inter-modality approaches, face recognition is performed on extracted discriminative features. Such features are color and texture descriptors which are invariant

to photo and sketch modalities. In comparison, inter-modality approaches are relatively new and therefore the majority of existing works synthesize pseudo photos from input sketches (or vice versa) into the same modality, followed by face recognition. Most face sketch/photo synthesis algorithms have been heavily influenced by face hallucination techniques [WTG+14].

Intra-modality approaches can be further categorised into four subcategories categories: *Bayesian inference* methods, *subspace learning* methods, a combination of *Bayesian* and *subspace learning* methods and *sparse representation* methods. Figure 4 shows all the subcategories of face sketch recognition in a tree diagram. The rest of this section analyses in greater detail all these approaches.

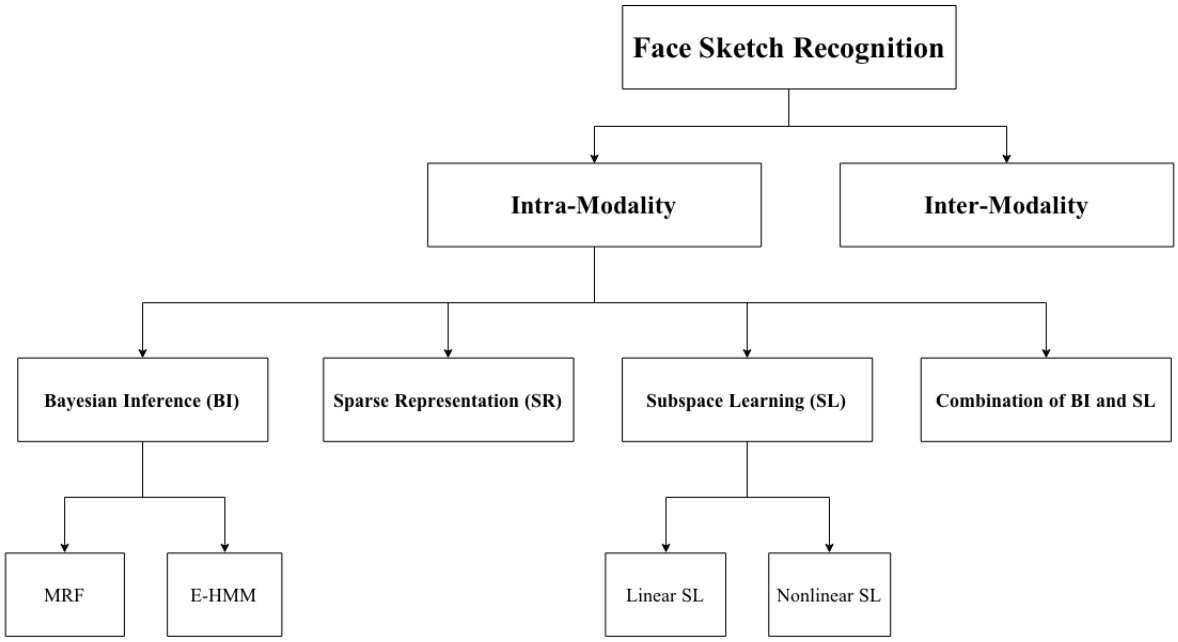

Figure 4: A tree diagram of the different categories of face sketch synthesis

3.1 Intra-modality Approaches

3.1.1 Bayesian Inference Method

The Bayesian inference method has presented itself into several sub-categories. Nonetheless, all methods follow a certain framework where prior probabilities $P(\theta)$ and likelihood terms $P(x|\theta)$ are calculated in order to derive the posterior probabilities. The major difference in each sub-category however is how these prior probabilities and the likelihood values

are calculated. The most imporatant sub-categories which are relevant to sketch/photo synthesis are *Markov random fields* and *embedded hidden Markov models*.

3.1.1.1 Markov Random Fields-based Method

The background section illustrated the manner in which several image processing and analysis problems can be described as labelling problems and how the model in Figure 3 can be used to describe the relationship between image patches. Freeman et al. [FPC00] proposed a framework where images and scenes (estimate output) were modelled by Markov Random Fields. Images were divided into overlapping patches and assigned to a node in a Markov network. For each input patch, N nearest neighbours and K candidates were found from the training set. The smoothness constraint ψ and compatibility ϕ relationships were constructed in order to calculate the joint probability. The local maximum *a posteriori* probability for each output patch was found through Bayesian belief propagation [YFW+00]. Patches were then combined by finding the average of the overlapping regions. Although the method in [FPC00] was applied to face super resolution, their work inspired many others to apply the same technique for sketch/photo synthesis.

Other related works [CXS+01][EL99][LSF07][LLX+01] showed promising results when adopting a patch-based nonparametric sampling technique for texture synthesis. These works inspired Liu et al. [LSZ01] to propose a non-parametric MRF-based super resolution method. Rather than model the likelihood and prior probabilities respectively they used PCA (Principal Component Analysis) to construct a global model that would create a global face image. The uniform scale of Markov random fields however is limited to just local dependency of patches. While this might yield more accurate results, since prominent features can be compared directly in each patch, one problem of this approach is that patches are only synthesized locally, thus making it more difficult to learn at a large scale, particularly the whole face structure.

Wang and Tang tackled this problem by using a multiscale MRF model where, from a training set, the photo-sketch model is learned at different patch sizes [WT09]. Inspired by Freeman et al. in [FPC00], candidate patches are chosen and assigned a node from the Markov model as in Figure 3 where through belief propagation, a synthesized image is produced. Other methods usually assume that all image patches have the same size. The drawback of this assumption is that an artist usually refers to the whole face structure when drawing a certain face area. However, using large fixed patch sizes tends to lead to more distortions and create a certain mosaic effect. As a result the multiscale MRF model learns

face structures at different scales in order to overcome these conflicts. Another advantage of multiscale MRF is that the approach can easily be used to either generate pseudo-sketches or pseudo-photos. Rather than finding the average of overlapping patches, image quilting [EF01] was used to reduce blurring effects between stitched patches. Their experiments showed that their proposed method achieved better visual results than previous methods.

Zhou et al. followed up on Wang and Tang's method by introducing what they call *Markov weight fields* [ZKW12]. They argue that Wang and Tang's method, although achieving optimal performance despite pose variations and illumination changes, can only synthesize an approximate solution and selecting just the best candidate makes it harder to synthesize unseen patches. They also slightly deviated from the Markov structure in Figure 3. The difference was that each node in the lower layer correspond to a list of variables or weights for candidate sketch patches, rather than to a single variable.

3.1.1.2 Embedded Hidden Markov Model

A hidden Markov model (HMM) is a means of representing probability distributions over sequences of observations [Gha01]. In any Markov model, the next state to be visited is chosen according to the state's transition probability distribution. The model hence produces two strings of information: i) the state path and ii) the observation sequence. The state path is a Markov chain which means that the next state does not depend on the sequence of events that come before, but rather on just the current state. If only an observation sequence is given, then the state path is a hidden Markov chain. In typical HMM problems one is required to find the best state path.

HMM are mostly used in speech recognition systems and other pattern recognition scenarios. Nonetheless, they have been successfully applied to computer vision problems as well. Samaria [Bal05] constructed a one-dimensional HMM where the face was partitioned into 5 regions: forehead, eyes, nose, mouth and chin. Each region relate to a hidden state while image intensities are taken as observations. The problem was that conventional HMM is challenging for face images that contain two-dimensional spatial information. Using conventional HMM would result in high computation and the loss of some spatial information. Embedded HMM was eventually proposed by Nefian and Hayes [NH99] and this drastically reduced computation cost.

E-HMM in this case consists of 5 super-states that model the face in a vertical manner. Each super-state corresponding to the forehead, eyes, nose, mouth and chin. Super-states are further divided into embedded states that describe that region in a horizontal manner.

Each super state acts as a one dimensional HMM. Gao et al. [GZLT08] used E-HMM to model the non-linear relationship between a sketch and its image pair. A series of pseudo-sketches were generated from several learned models and fused together with selective ensemble strategy in order to generate a finer pseudo-sketch.

The method introduced in [GZLT08] was only conducted on a holistic face image, hence some noise was introduced since certain fine features associated with the eyes, nose and mouth could not be learned. Thus, Gao et al. extended their approach to local patch-based sketch synthesis [GZTL08]. Images were divided into overlapping patches where for each input patch the target patch is synthesized using the same approach introduced in [GZLT08]. Xiao et al. [XGTL09] applied Gao et al.'s method for face photo synthesis.

3.1.2 The Subspace Learning Framework

The Subspace learning technique can be described as finding a subspace $\mathbb{R}^m$ within a higher dimensional space $\mathbb{R}^n$ such that $n > m$. Principle component analysis (PCA) and linear discriminant analysis (LDA) are currently two of the widely used subspace learning techniques, which are mainly used for face recognition. The main idea of these techniques is that the face space has a lower dimension than the image space [Niy04]. This involves constructing a projection matrix $U \in \mathbb{R}^{n \times m}$ which is learned from the training examples. The projection matrix can be calculated by using standard eigenvalue decomposition [ZTLY09] or generalized eigenvalue decomposition [Niy04]. There are two sections for the subspace learning framework: linear subspace learning and nonlinear manifold learning.

Tang and Wang [TW02], [TW03], [TW04] proposed an eigentransformation method for face sketch synthesis by using PCA. In PCA reconstruction a new face sketch (or photo) can be described by a linear combination of photo-sketch pair samples. Eigentransformation computes the weighted combination of sketch eigenfaces to reconstruct the sketch training image. these derived weights are then used to combine the set of photo images to derive the pseudo-image. If T_p and T_s are the photo and sketch pairs respectively, the formulae $I_p = T_p C_p$ and $I_s = T_s C_s$ represents this linear combination where each column of T_p and T_s represent corresponding training samples while C_p and C_s are column vectors obtained through PCA eigentransformation. The assumption is that sketch reconstruction is similar to its image pair, therefore $C_p \approx C_s$.

Since sketches do not obtain optimal face recognition performance, Li et al [LSB06] introduced an algorithm for synthesizing a photo from its corresponding sketch. They used eigen-analysis on a hybrid space of training sketches and training photos, rather than

just on the photo space.

Approximating images through a linear process might turn out to be quite inaccurate. Liu et al. [LTJ+05] introduced two nonlinear techniques for synthesizing images. Inspired by local linear embedding (LLE), mapping the relation between photos and sketches could be achieved by using a patch based strategy. LLE is a learning method that approximates non linearity by a local linearity. In other words, it is a method where weights are computed for each patch through a linear combination of its neighbours. Their experimental results showed that this nonlinear method achieved better results than in [TW02][TW03] and [TW04]. The rest of the research in this sub category lead to further development in super resolution techniques.

3.1.3 Combination of Bayesian Inference and Subspace Learning Framework

The approach adopted by Liu et al. [LSF07][LSZ01], although categorised under the Bayesian inference section, can also be deemed to be a representative technique of the Bayesian inference and subspace learning category [WTG+14]. In their approach PCA was used to obtain an initial global face image and then proceeded to use MAP-MRF to calculate local face images.

Later on, Liu et al introduced a two-step procedure for photo synthesis from an input sketch [LTL07]. The first step was inspired by [LTJ+05] (LLE-based) in order to generate an initial estimate. Through a proposed tensor model whose modes consist of patch position, style, and features, the high frequency residual error is inferred under the MAP framework. It is assumed however that sketch-photo pairs have the same tensor representation parameter. The addition of these two steps could synthesize a detailed photo.

3.1.4 The Sparse Representation-based Approaches

Sparse representation tries to find the sparse weighted combination of a set of training sketches (referred to as atoms [WTG+14]) that can reconstruct the test sketch image. The same sparse weights are used to combine the corresponding photo training photos. Initially Yang et al [YTMH08] proposed a face super resolution method based on sparse coding which eventually inspired Chang et al [CZHD10] to use the same model for sketch-photo synthesis. The first step was to construct a coupled dictionary of training photo and sketch pairs by using sparse coding. Next, for each patch in the test photo, its sparse representation is obtained as a product of the sketch patch and its sparse representation

coefficient. The sketch image is finally reconstructed by enforcing a smoothness constraint between overlapping patches.

3.2 Inter-modality Approaches

Recently a few researchers have instead focused on minimizing the modality difference when extracting features. Klare and Jain [KJ10] proposed the first inter-modality method. Dense scale-invariant feature transform (SIFT) descriptors [Low99] are extracted from patches in order to reconstruct a holistic image representation. A simple 1-NN classifier is then used for sketch-photo matching. Klare et al. [KLJ11] went further and proposed local feature based discriminant analysis (LFDA) to match sketches to photos. Here sketches and photos are represented by two features: SIFT descriptors and multi local binary patterns (MLBP).

LFDA however was not enough to minimize the difference between the two modalities. Zhang et al. [ZWT11] introduced a face descriptor based on coupled information-theoretic encoding to extract modality-invariant descriptor. Coupled information-theoretic projection increases the mutual information between photos and sketches. Their method is considered as state of the art [GS12].

Another new gradient orientations based face descriptor was introduced by Galoogahi et al. [GS12]. This was called Histogram of Averaged Oriented Gradients (HAOG). with HAOG, the modality gap between sketches and photos is reduced drastically by emphasizing coarse texture of facial components when extracting features.

4 Specification and Design

In this dissertation, an intra-modality approach was used for face sketch recognition. The main focus was to research and implement a face image synthesis algorithm using Markov Random Fields. This algorithm was applied to a specific scenario where law enforcement must rely on a sketch based on witness recollection in order to identify the suspect. Essentially one can reach the solution in two different ways:

1. By comparing the query sketch with synthesized pseudo-sketches of the photos within the database/gallery or

2. By comparing a synthesized pseudo-photo of the query sketch with real photos in the database/gallery.

The proposed synthesis algorithm is not tied to one specific conversion, hence the algorithm works both ways. Clearly the second would seem to be the most practical for this scenario. First of all, most existing face recognition systems at the moment are designed to match photo based faces, not sketches [XGTL09]. Also a lot of pre-processing would be required in order to generate pseudo-sketches for every photo in the photo database. As a result it would be cheaper and quicker to generate just one pseudo-photo from the query sketch and use existing algorithms for face recognition. On the other hand, some literature suggest that the conversion of photos to sketches is more accurate for recognition. The reason is that since photos are richer in detail, converting photos to sketches would actually involve reduction of information. Nonetheless, both methods were evaluated in this dissertation.

A high level overview of the proposed algorithm can be seen in Figure 5. A training set containing photo-sketch pairs is required with the assumption that image pairs are similar in shape and distinct features. These training images are warped into the same face template as show in Figure 5. The scope is to align training images by transforming them into one specific face template. The input sketch is also warped in the same face template. The input warped sketch is then divided into equal patches. K candidate patches pairs are then chosen from the training set which best match these input patches. These candidate patches are then inserted into a Markov model such that each node corresponds to a patch. The Markov model defines the relationship between query, candidate and neighbouring candidate patches. Using Markov random fields, the best patch is chosen from the candidate training patches. This patch must not only be similar to the query patch, but it must also fit well with neighbouring patches. Finally overlapping patches are stitched back together by averaging overlapping pixels in order to form an image.

There are many mathematical ways in which the relationship between patches can be modelled such as kNN based algorithms and other Markov Models. However, the proposed methodology uses Markov random fields which is suitable for modelling the *a priori* probability of context dependent patterns including object features [Li09]. Another reason for using MRF was that in literature the MRF approach seemed to produce the most promising results, especially Wang and Tang's state of the art method Multiscale MRF Model [WT09]. In MRF theory, a patch can receive information not just from neighbouring patches, but also from patches further away through belief propagation. This however would severely increase complexity and it was decided to stick to just a first order neighbourhood system just like in Figure 2(a).

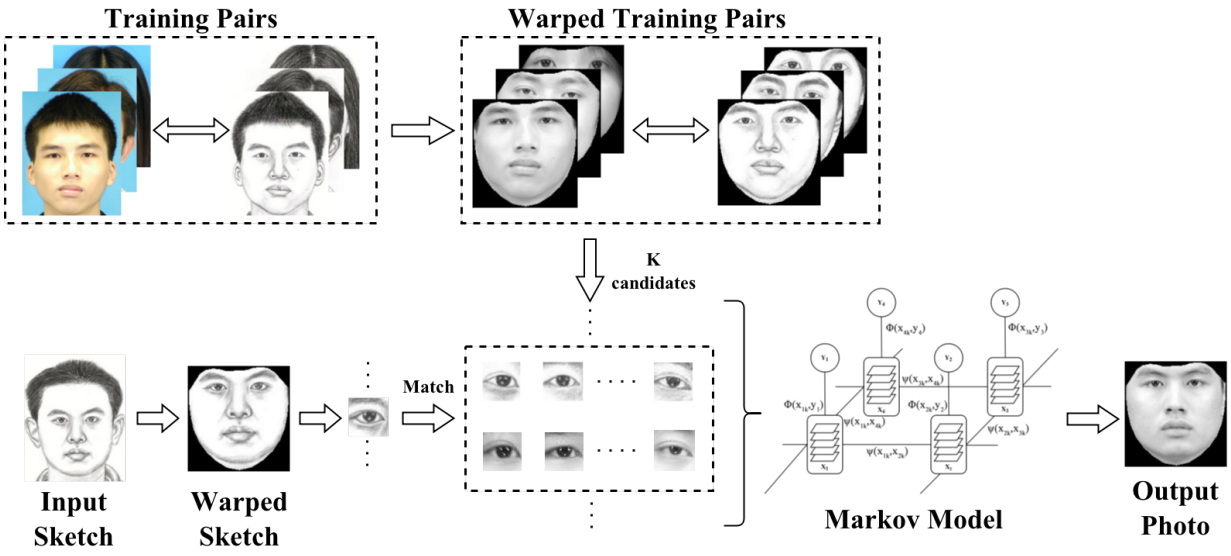

Figure 5: A high level system view of the algorithm

5 Implementation

Throughout this section the main focus will be on face photo synthesis, namely constructing a pseudo-photo from an input face sketch. The approach can easily be extended to sketch synthesis by substituting roles of photos and sketches. The steps required for face photo synthesis are explained sequentially in the following subsections. A training set of photo-sketch pairs is required with the general assumption that photo-sketch pairs are recognizably the same person.

5.1 Pre-processing

The first step was to scale and align all images so that different face components roughly lie in the same face region. There are several ways to align face images however it was decided to use some functionality from the Active Orientation Models (AOMs) introduced by Georgios, et al. [TAiMZP14]. AOMs are generative models of facial shape and appearance, an extension of the Active Appearance Models (AAMs) which is a well-known computer vision algorithm for matching statistical models of appearance to images. Georgios, et al. have kindly released their MATLAB code for public use. The appearance model of an AOM is learned by first warping training images to a canonical reference frame. The exact same face warping functions were used in this dissertation in order to transform all face images into the same reference frame.

All photos were converted to grayscale. Converting to other colour spaces such as the Luv or Lab colour space is also an option. Some researchers claim that Euclidean distance would perform better with Luv colour space however grayscale still achieved optimal results.

5.2 Patch Matching

The next step was to divide the warped face images into overlapping patches. Simply dividing patches equally across all images would not be ideal because although faces have been roughly aligned in the previous step, certain face features vary in shape, size and relative position and will still not be perfectly aligned. To help solve this problem, the input face sketch was first divided into equal overlapping patches. These patches were used as reference patches. For each reference patch its corresponding position in all training photos was located. This required comparing the reference patch to all possible patches in the training image, however rather than searching the whole image, a search radius was defined. The best matching patch was found using simple Euclidean distance. Out of these training patches, the K candidate patches which best matched the corresponding reference patch were chosen. This was done by using a Structural Similarity Index Metric (SSIM) algorithm that computes a similarity score based on luminance, contrast and structure. The greater the score, the more similar the patches are to one another. Since the query patch is a sketch patch, SSIM indexing was performed on training sketch patches. This process was followed by the selection of the nearest K candidate patches and their corresponding photo patch pair. The reason for choosing K candidate patches is that patch estimation in this algorithm takes into consideration how well a patch fits in with its neighbouring patches. If only one candidate patch were to be chosen, then the synthesized photo would most likely have a mosaic effect with disoriented face features. While the synthesized photo should match the input sketch as closely as possible, at the same time the image must be smooth in appearance. In order to reach this goal, a Markov network was used to model the process of face image synthesis.

5.3 Markov network

Figure 6 shows the Markov model that was adopted for this particular problem. Since images were divided into overlapping patches, each patch was assigned a node. Let y_i and x_i represent the input sketch patch and the estimate photo patch at patch i respectively.

As previously mentioned, candidate training patches that best match patch y_i were chosen, hence node x_i in the model also contained candidate patches, each with its own notation. Candidate patch k at patch i is denoted as x_{ik}.

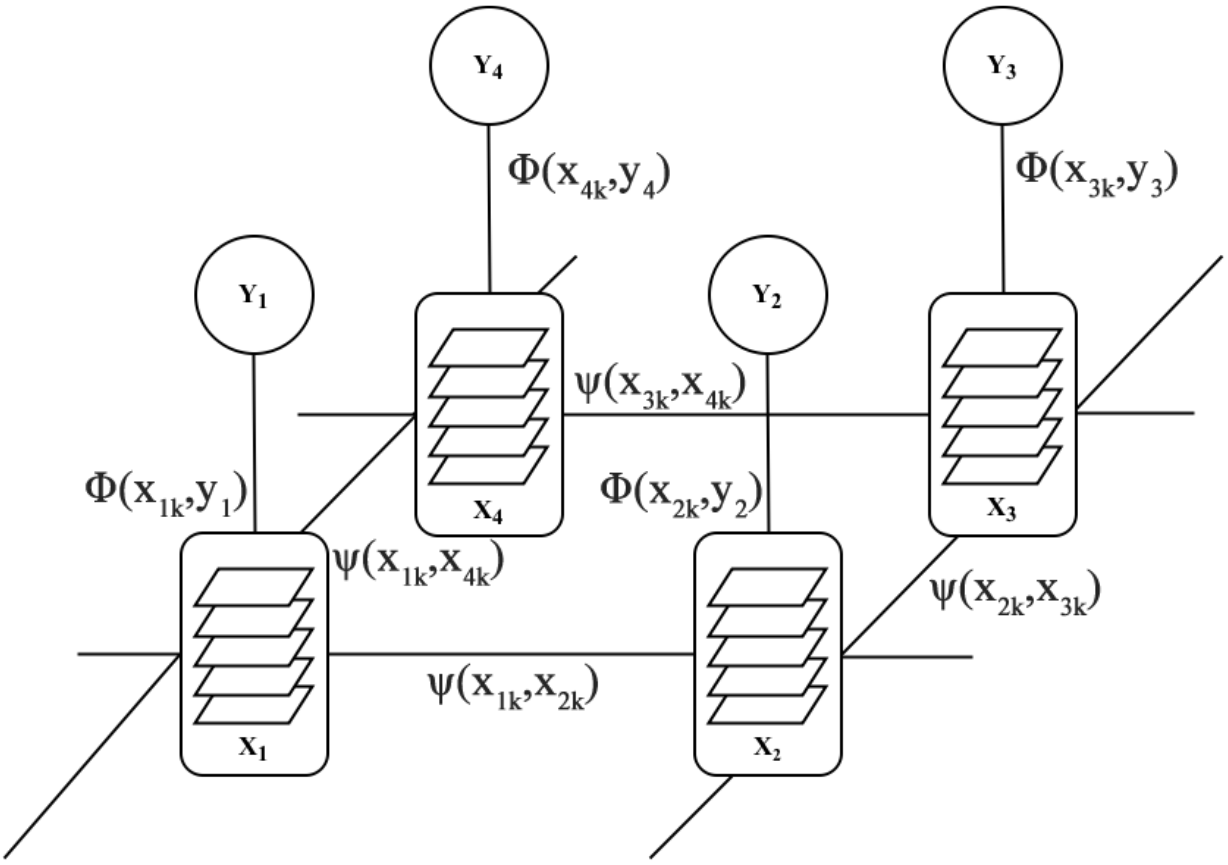

Figure 6: The graphical model of the Markov network that will be used. Y and X represent input sketch patches and candidate estimate patches respectively.

The relationship between patches was defined by the compatibility functions ψ and ϕ. The dependency between x and y is denoted by $\phi(x_i, y_i)$ which provides local evidence for x_i. Estimate patches x_i on the other hand are connected to neighbouring estimate patches with the function $\psi(x_i, x_j)$ where x_j is neighbouring node. The dependency between x_i and y_i is calculated as:

$$\phi(\tilde{x}_i, y_i) = \exp\{-\|\tilde{y}_i - y_i\|^2 / 2\sigma_e^2\} \tag{8}$$

where $\tilde{x}_i$ refers to the estimated photo patch, $\tilde{y}_i$ its corresponding sketch patch pair, y_i the input reference patch and σ_e the variance that needs to be tuned empirically. It is assumed that $\tilde{x}_i$ and $\tilde{y}_i$ are very similar in structure. All patches overlap each other by a certain distance parameter. Let x_{ik} and x_{jk} be two overlapping patches. The overlapping region

for x_{ik} and x_{jk} is denoted as d_{ij}^{l} and d_{ji}^{m} respectively as in Figure 6. These overlapping regions are used in order to compute the compatibility matrix between patches i and j as:

$$\psi(x_{ik}^{l}, x_{jk}^{m}) = \exp\{-\|d_{ij}^{l} - d_{ji}^{m}\|^2 / 2\sigma_e^2\} \tag{9}$$

By using these dependency and compatibility functions, the joint probability between the input sketch and the synthesized photo can be written as:

$$P(x_1, x_2, ..., x_N, y_1, y_2, ...y_N) = \prod_{(i,j,k)} \psi(x_{ik}, x_{jk}) \prod_{l} \phi(x_l, y_l) \tag{10}$$

By taking the maximum *a posteriori* (MAP) estimator $\hat{x}_{iMAP}$, we could estimate the synthesized patch from the constructed Markov network:

$$\hat{x}_{i\,MAP} = \arg\max_{x_i} \arg \max_{[all\, x_j, i \neq j]} P(x_1, x_2, ..., x_N, y_1, y_2, ...y_N) \tag{11}$$

Computing the MAP estimates is the most time consuming part of the algorithm. After calculating all dependency and compatibility probability values, for every candidate patch it is required to multiply its dependency value with all possible compatibility values from neighbouring candidate patches. This involves computing the probability score for all possible combinations of patch match. The candidate patch which best fits neighbouring patches and its corresponding reference patch would have the highest probability score, and would be chosen as the estimate patch. Calculating the MAP estimator can be extended along any neighbourhood radius, however it was decided to work with just first order neighbourhood system, namely with the closest four neighbours. Extending beyond four neighbours would have increased the complexity even further.

Execution time highly depends on the patch size, overlap size and number of candidates in this case. Smaller patch and overlap sizes drastically increase the number of patches while the number of candidates affects the number of combinations that must be processed.

5.4 Stitching Patches

When the output photo patches have been estimated, all that remained was to reconstruct an image from these overlapping patches. Within this algorithm, the average of overlapping pixels was calculated. The output image was finally ready for face recognition.

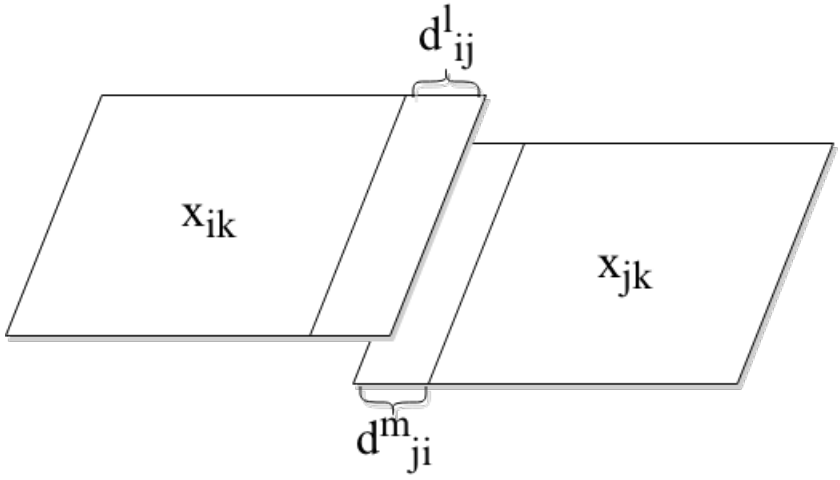

Figure 7: The compatibility between two patches is determined by the values in their region of overlap.

5.5 Other Remarks

Unlike most photo/sketch synthesis approaches, hair region has been ignored when synthesizing an image. This is because the presence of hair does not affect the recognition performance of modern face recognizers [TKB12].

6 Testing and Evaluation

The implemented algorithm was tested using the CUHK database [WT09] which consists of 188 photo-sketch pairs. Tests were carried out for both sketch to photo and photo to sketch synthesis. These tests can be categorised into two sections: quality tests and recognition tests.

6.1 Quality Tests

In this section 88 image pairs were used for training while the remaining 100 images were used for testing as in [WT09]. The quality of the resulting synthesized image can be described by two types of measurements: peak signal-to-noise ratio (PSNR) and the structural similarity index metric(SSIM). PSNR is measured in decibels (dB) and is a valid quality measure [HTG08]. On the other hand SSIM is a technique for measuring similarity based on luminance, contrast and structure. SSIM is known to be more correlated to the

human perception. For both measurements, the synthesized photo is compared to the actual photo and vice-versa. High PSNR and SSIM scores signify high similarity.

There are many parameters which can influence the final result namely patch size, overlap size, number of candidates, processing image size and search space radius. However it was decided that it would be sufficient to focus on different patch sizes and number of candidates.

6.1.1 Patch Sizes

All images were resized to 160 $\times$ 160 pixels before processing. Experimentation was done with patch sizes of 5, 10, 15 and 20 pixels. The overlap size was taken as half the patch size while 5 candidates were extracted from the training set. The results for patch size performance can be seen in Figure 8. For both metrics the y-axis consists of the cumulative probability. For example, for a PSNR of xdB the cumulative probability gives the probability that the quality is lower than x.

Clearly the ideal patch size is 20 pixels. This was quite expected since with 20 pixels there are less patch overlaps. Since the average is taken between overlapping pixels, overlapping regions are susceptible to noise. A surprising result was that patch size 10 outperformed patch size 15. Since 15 is an odd number, the overlap was set to 7 pixels. A plausible reason for this result could be because of the extra pixels that are not being overlapped in odd sized patches. These pixels probably do not fit in smoothly with neighbouring overlapped pixels, therefore increase noise slightly.

Another expected observation is that overall, photo to sketch synthesis performed better. This was expected because the conversion of photos to sketches essentially consists of reducing information which is much easier to do.

6.1.2 Number of Candidates

Choosing K candidate patches from the training set is an essential part of the methodology since choosing the right patch includes computing the maximum *a posteriori* estimate from a set of neighbouring candidate patches. Since there are 88 training images, the maximum number of candidates is 88. The results are as shown in Figure 9. Given the results in the previous tests, the following candidate tests were run with 20 pixels patch size and 10 pixel overlap.

The first question to ask is whether taking candidates is necessary at all (therefore $K > 1$), hence only taking the closest patch into consideration. If only one candidate was

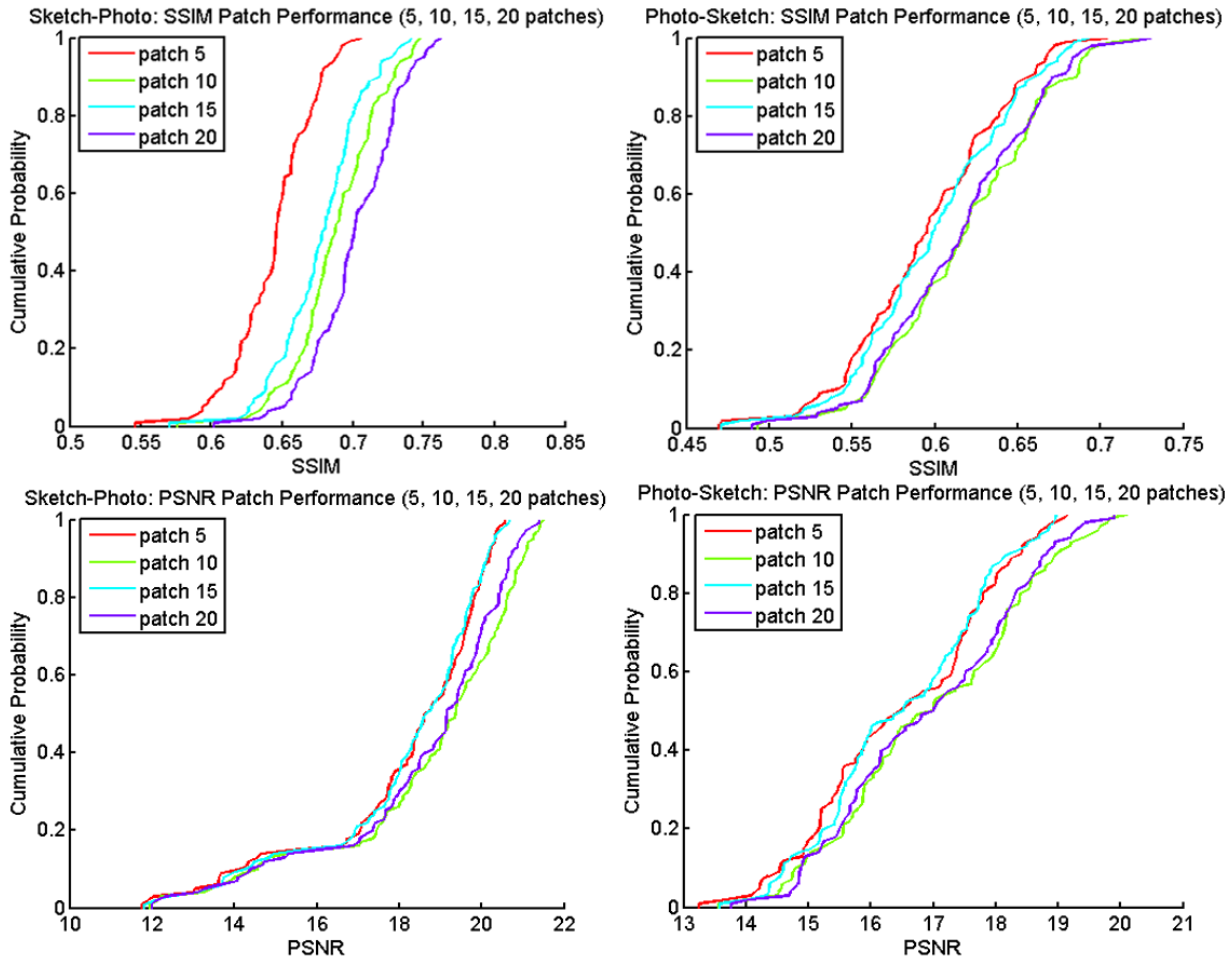

Figure 8: Patch performance for sketch-photo synthesis and photo-sketch synthesis measured in PSNR and SSIM. The further the curve is to the right, the better the performance.

needed, then calculating the MAP estimate, particularly with the smoothness constraint ψ, would be redundant. From the results it was evident that more than one candidate was needed to achieve optimal results. This result was expected since previous research has shown that compatibility with neighbouring patches must also be taken into consideration.

The next question to ask is how many candidates are actually needed. Clearly taking all patches as candidates gave the worst results. This seems logical because patches that received a low similarity score to the reference patch might still achieve a high compatibility score with neighbouring patches. This result highlights the importance of filtering out patches as candidates. The fact that only the top K matching patches are taken as candidates gives a certain priority on reference patch similarity. The ideal number of candidates

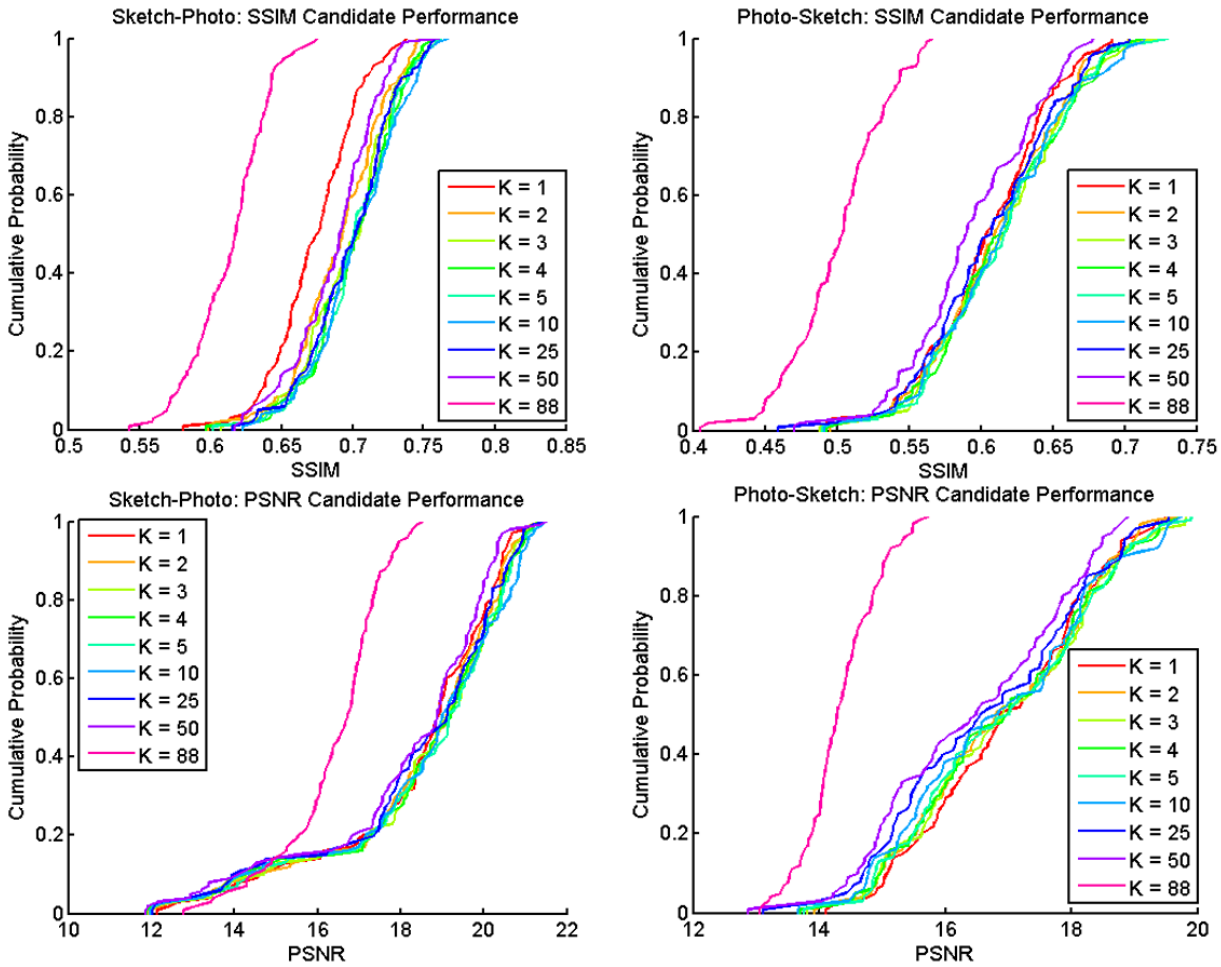

Figure 9: Candidate performance for sketch-photo synthesis and photo-sketch synthesis measured in PSNR and SSIM. The further the curve is to the right, the better the performance.

seems to lie between 3 to 5 candidates, depending on the type of synthesis. There was no particular expected number of candidates, but the resulting range is acceptable. Again, photo to sketch synthesis performed much better than sketch to photo synthesis.

6.2 Recognition Tests

For these tests, the leave-one-out technique was used for synthesizing images. This means that for every input image, the remaining images were used as training. Using the optimal parameters found in the previous quality tests, Figure 10 shows some examples of both sketch and photo synthesis results.

For these tests, the PHD (Pretty Helpful Development) toolbox [ŠP09][ŠP10] was used.

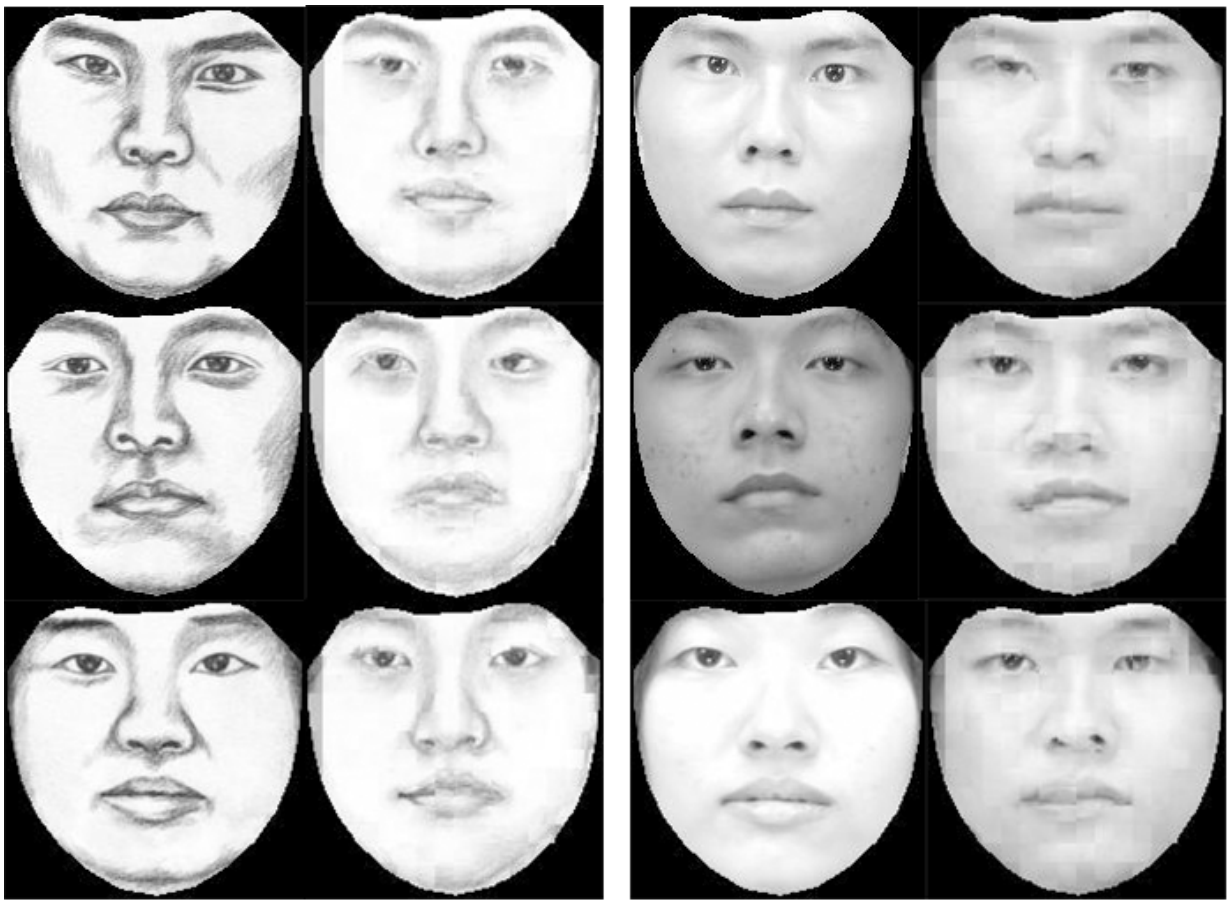

Figure 10: Face sketch synthesis and face photo synthesis results.

The PHD toolbox consists of several MATLAB implementations for popular face recognition techniques such as PCA, LDA and Gabor filtering along with some useful plotting functions. The PCA face recognition technique was used in these recognition tests since it is commonly used in literature. LDA is also a popular technique however it requires more than one training image per subject.

It is pertinent to recall the scenario where police officers must rely on a sketch drawn by a forensic artist. The tests revolved around the two approaches of preparing sketches for automatic face recognition: either generating a synthesized pseudo-photo from the input sketch, or synthesizing all database images to pseudo-sketches. Although existing face recognizers are designed to match photo based faces, a face recognizer can still recognize sketches if it was trained on a sketch database. A total of 4 tests were performed. The base test consisted of feeding a photo trained face recognizer with input sketches. Two

other experiments test the aforementioned methods of preparing a sketch for automatic face recognition. Given the very positive results obtained from training the recognizer with synthesized sketches, it was also decided to test whether doing the complete opposite, therefore training the recognizer with the original sketches and testing with synthesized sketches, obtained the same result. The recognition results are shown in Table 1 and Figure 11.

Table 1: Rank 1 recognition results for the three scenarios of face sketch recognition.

Type	**Rank 1 (in %)**	**Rank 10 (in %)**
Direct Sketch Recognition	35.64	62.30
Photo Synthesis	20.21	48.41
Sketch Synthesis (as Training)	94.68	98.23
Sketch Synthesis (as Testing)	86.70	99.01

Figure 11 illustrates the Cumulative Match Characteristic (CMC) curve for each scenario. In face recognition systems, a rank is used to measure face recognition performance. The recognition rate at rank N refers to the success rate of identifying the correct face image from a set of N nearest candidates that are chosen by the face recogniser as the closest match. The lower the rank, the harder it is to obtain a high recognition score.

The biggest question that needs to be answered by these recognition tests is whether converting to the same modality is necessary for improving recognition results. While from Table 1 poor results were achieved when converting sketches to photos, the opposite proved otherwise. Given the quality test results, it comes to no surprise that sketch synthesis performed better than photo synthesis. The main reason for poor results for photo synthesis is that photo synthesis is much harder and more prone to noise than sketch synthesis since a lot of texture detail is involved when dealing with photo patches. Another reason is that it is hard to get the illumination information from a sketch. PCA is known to be susceptible to lumination variation. A minor solution would be smoothing overlapping patches through image quilting [EF01] in order to reduce the mosaic effect that is evident in Figure 10. However the biggest improvement would be to have a much larger training set. Another interesting detail was that although better recognition results were achieved when training the recognizer with synthesized sketches rather than with sketches, 100% recognition rate was achieved at a much lower rank when training with sketches.

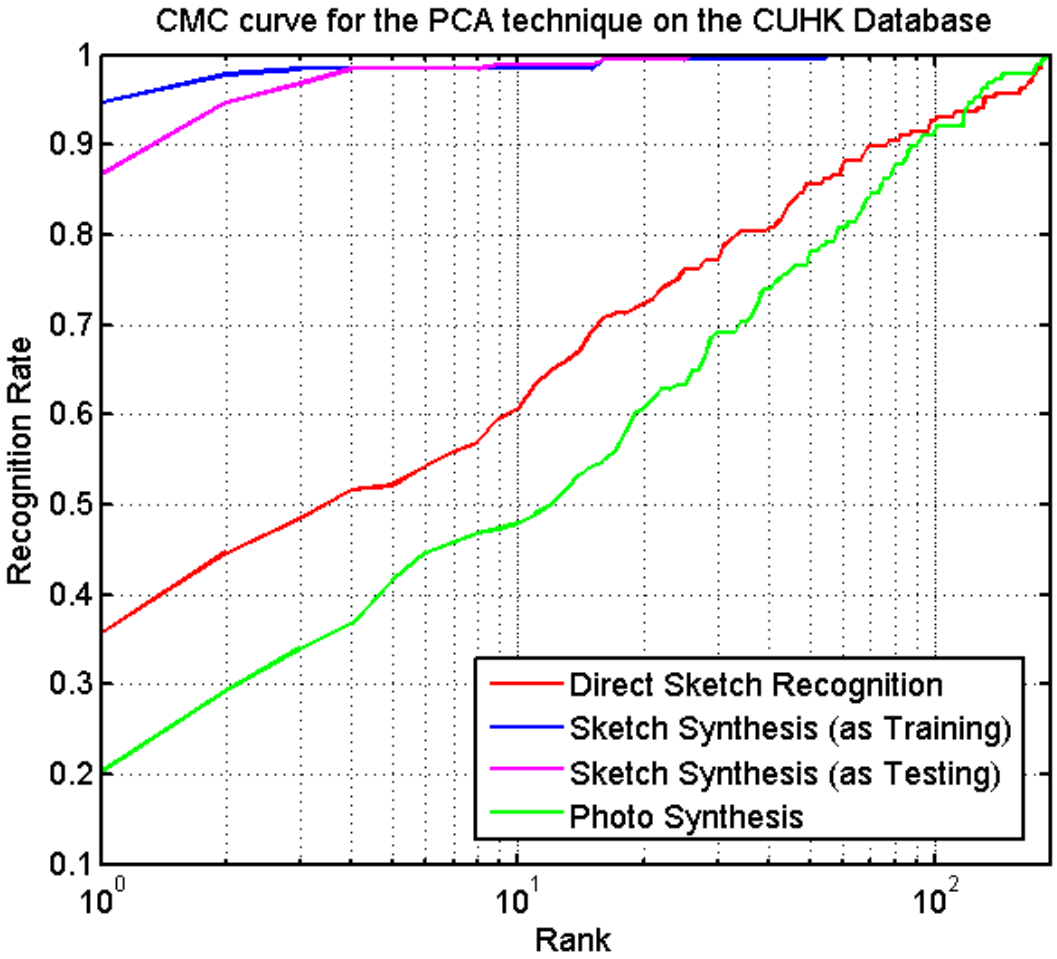

Figure 11: PCA recognition results.

7 Conclusions and Future Work

In this dissertation, a face sketch/photo synthesis algorithm based on Markov random fields was implemented in order to prepare sketches for automatic face recognition systems. The aim of this dissertation was to investigate whether converting images to the same modality improves the face recognition rate. Although this was not so evident for photo synthesis, it was more successful in the case of sketch synthesis. Both quality tests and recognition tests have shown that sketch synthesis yielded better recognition results than photo synthesis, primarily because sketch synthesis involves reducing information. Other observations have shown that 20 pixels is the optimal patch size, when dealing with 160×160 images, while the ideal number of candidates is between 3 and 5 candidates.

Despite these results, there is still a lot of room for improvement, especially when it

comes to synthesizing photos. One observation for improving photo synthesis would be to obtain a larger training set. Compared to sketches, skin texture within photos contains much more information such as different skin tone, shadowing, and feature details. More image information results in more possible patch combinations. Therefore having more patches to choose from would be considered an asset.

Future work could include smoothing overlapping patches through image quilting [EF01] rather than simply taking their average. This would reduce the evident blurring effect by making a minimum error boundary cut between overlapping patches. This blurring effect is quite evident in pseudo-photos since different patches will have different skin tones and textures.

Another problem of this approach is the blind assumption that is made between photo and sketch pairs. For example, when synthesizing a photo, the top K training patches that best match the input sketch patch are chosen. However their patch pair is used to generate the resulting image. Currently the relationship between training image patch pairs are defined by their location in the image, yet there is no guarantee that local sketch patch neighbourhoods are preserved in photo space. Typically in pencil drawings, certain structures and face features might be slightly exaggerated in comparison with the real photo. Using the candidate pairs with this blind assumption would eventually lead to distorted face features. This is quite evident in Figure 10 particularly in the mouth, nose and eye regions. Recently Bevilacqua et al. [BRGM13] sought to enforce the coherence between low resolution and high resolution patches. They proposed a two-step dictionary learning strategy. First, a clustering process gathers the low resolution and high resolution patches into jointly coherent clusters. This is then followed by extracting a set of particularly representative patches that can express the whole dictionary in a compact way from the clustered dictionary. This might also be applicable to sketch/photo synthesis problems.

Finally future work should also be algorithm optimisation. On a computer with a 3GHz CPU, depending on the parameters, image processing could take between 3 to 20 minutes per image. The algorithm was implemented in a serial fashion and could easily be optimised to use make use of more CPU or GPU cores given that each patch can be processed independently.

Acknowledgements.

I would like to thank my tutor Dr. Ing. Reuben A. Farrugia for his constant guidance, input and support throughout this year. His understanding and experience has helped me immensely.

I would also like to thank all the lecturers in the Faculty who in their own way helped me throughout the course.

I would also like to thank my close friends whose friendship and laughter made my university years a memorable time.

Last but not least, I would like to thank my family for their constant support and encouragement.

Appendices

All code was implemented in MATLAB and can be found on the CD. The CD also contain results obtained from the quality and recognition tests. A README file is included with further details about the contents of the CD.

References

[Awa07] Suyash P. Awate. Maximum-a-posteriori (map) estimation. `http://www.cs.utah.edu/~suyash/Dissertation_html/node8.html`, 2007.

[Bal05] Johan Stephen Simeon Ballot. *Face recognition using Hidden Markov Models*. PhD thesis, Stellenbosch: University of Stellenbosch, 2005.

[Bay] Basics of bayesian inference and belief networks. `http://research.microsoft.com/en-us/um/redmond/groups/adapt/msbnx/msbnx/Basics_of_Bayesian_Inference.htm`.

[BHD+92] Vicki Bruce, Elias Hanna, Neal Dench, Pat Healey, and Mike Burton. The importance of mass in line drawings of faces. *Applied Cognitive Psychology*, 6(7):619–628, 1992.

[BP91] Philip J Benson and David I Perrett. Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, 3(1):105–135, 1991.

[BRGM13] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and M-LA Morel. Compact and coherent dictionary construction for example-based super-resolution. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 2222–2226. IEEE, 2013.

[CET01] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001.

[CXS+01] Hong Chen, Ying-Qing Xu, Heung-Yeung Shum, Song-Chun Zhu, and Nan-Ning Zheng. Example-based facial sketch generation with non-parametric sampling. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 433–438. IEEE, 2001.

[CZHD10] Liang Chang, Mingquan Zhou, Yanjun Han, and Xiaoming Deng. Face sketch synthesis via sparse representation. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2146–2149. IEEE, 2010.

[EF01] Alexei A Efros and William T Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 341–346. ACM, 2001.

[EL99] Alexei A Efros and Thomas K Leung. Texture synthesis by non-parametric sampling. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1033–1038. IEEE, 1999.

[FPC00] William T Freeman, Egon C Pasztor, and Owen T Carmichael. Learning low-level vision. *International journal of computer vision*, 40(1):25–47, 2000.

[FTP99] William T Freeman, Joshua B Tenenbaum, and Egon Pasztor. An example-based approach to style translation for line drawings. *Mitsubishi Elect. Res. Lab., Cambridge, MA, MERL Tech. Rep. TR99-11*, 1999.

[Gha01] Zoubin Ghahramani. An introduction to hidden markov models and bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(01):9–42, 2001.

[GS12] Hamed Kiani Galoogahi and Terence Sim. Inter-modality face sketch recognition. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, pages 224–229. IEEE, 2012.

[GZLT08] Xinbo Gao, Juanjuan Zhong, Jie Li, and Chunna Tian. Face sketch synthesis algorithm based on e-hmm and selective ensemble. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(4):487–496, 2008.

[GZTL08] Xinbo Gao, Juanjuan Zhong, Dacheng Tao, and Xuelong Li. Local face sketch synthesis learning. *Neurocomputing*, 71(10):1921–1930, 2008.

[HTG08] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008.

[ITO99] Shino Iwashita, Yyyju Takeda, and Takehisa Onisawa. Expressive facial caricature drawing. In *Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE'99. 1999 IEEE International*, volume 3, pages 1597–1602. IEEE, 1999.

[KJ10] Brendan Klare and Anil K Jain. Sketch-to-photo matching: a feature-based approach. In *SPIE Defense, Security, and Sensing*, pages 766702–766702. International Society for Optics and Photonics, 2010.

[KLJ11] Brendan F Klare, Zhifeng Li, and Anil K Jain. Matching forensic sketches to mug shot photos. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(3):639–646, 2011.

[KTFM99] Hiroyasu Koshimizu, Masafumi Tominaga, Takayuki Fujiwara, and Kazuhito Murakami. On kansei facial image processing for computerized facial caricaturing system picasso. In *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on*, volume 6, pages 294–299. IEEE, 1999.

[Li09] Stan Z Li. *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009.

[LLX+01] Lin Liang, Ce Liu, Ying-Qing Xu, Baining Guo, and Heung-Yeung Shum. Real-time texture synthesis by patch-based sampling. *ACM Transactions on Graphics (ToG)*, 20(3):127–150, 2001.

[Lov11] I. Lovett. Los angeles officials identify video assault suspects. `http://www.nytimes.com/2011/01/08/us/08disabled.html`, 2011.

[Low99] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[LSB06] Yung-hui Li, Marios Savvides, and Vijayakumar Bhagavatula. Illumination tolerant face recognition using a novel face from sketch synthesis approach and advanced correlation filters. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 2, pages II–II. IEEE, 2006.

[LSF07] Ce Liu, Heung-Yeung Shum, and William T Freeman. Face hallucination: Theory and practice. *International Journal of Computer Vision*, 75(1):115–134, 2007.

[LSZ01] Ce Liu, Heung-Yeung Shum, and Chang-Shui Zhang. A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–192. IEEE, 2001.

[LTJ+05] Qingshan Liu, Xiaoou Tang, Hongliang Jin, Hanqing Lu, and Songde Ma. A nonlinear approach for face sketch synthesis and recognition. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 1005–1010. IEEE, 2005.

[LTL07] Wei Liu, Xiaoou Tang, and Jianzhuang Liu. Bayesian tensor inference for sketch-based facial photo hallucination. In *IJCAI*, pages 2141–2146, 2007.

[NH99] A Nefian and M Hayes. Face recognition using an embedded hmm. In *IEEE Conference on Audio and Video-based Biometric Person Authentication*, pages 19–24, 1999.

[Niy04] X Niyogi. Locality preserving projections. In *Neural information processing systems*, volume 16, page 153. MIT, 2004.

[ŠP09] Vitomir Štruc and Nikola Pavešić. Gabor-based kernel partial-least-squares discrimination features for face recognition. *Informatica*, 20(1):115–138, 2009.

[ŠP10] Vitomir Štruc and Nikola Pavešić. The complete gabor-fisher classifier for robust face recognition. *EURASIP Journal on Advances in Signal Processing*, 2010:31, 2010.

[TAiMZP14] Georgios Tzimiropoulos, Joan Alabort-i Medina, Stefanos P Zafeiriou, and Maja Pantic. Active orientation models for face alignment in-the-wild. *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, 9(12), 2014.

[TKB12] Umar Toseeb, David RT Keeble, and Eleanor J Bryant. The significance of hair for face recognition. *PloS one*, 7(3):e34144, 2012.

[TW02] Xiaoou Tang and Xiaogang Wang. Face photo recognition using sketch. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–257. IEEE, 2002.

[TW03] Xiaoou Tang and Xiaogang Wang. Face sketch synthesis and recognition. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 687–694. IEEE, 2003.

[TW04] Xiaoou Tang and Xiaogang Wang. Face sketch recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):50–57, 2004.

[UJdVL96] Robert G Uhl Jr and Niels da Vitoria Lobo. A framework for recognizing a facial image from a police sketch. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, pages 586–593. IEEE, 1996.

[WT09] Xiaogang Wang and Xiaoou Tang. Face photo-sketch synthesis and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(11):1955–1967, 2009.

[WTG+14] Nannan Wang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Jie Li. A comprehensive survey to face hallucination. *International journal of computer vision*, 106(1):9–30, 2014.

[XGTL09] Bing Xiao, Xinbo Gao, Dacheng Tao, and Xuelong Li. A new approach for face recognition by sketches in photos. *Signal Processing*, 89(8):1576–1588, 2009.

[YFW+00] Jonathan S Yedidia, William T Freeman, Yair Weiss, et al. Generalized belief propagation. In *NIPS*, volume 13, pages 689–695, 2000.

[YTMH08] Jianchao Yang, Hao Tang, Yi Ma, and Thomas Huang. Face hallucination via sparse coding. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 1264–1267. IEEE, 2008.

[ZKW12] Hao Zhou, Zhanghui Kuang, and K-YK Wong. Markov weight fields for face sketch synthesis. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1091–1097. IEEE, 2012.

[ZTLY09] Tianhao Zhang, Dacheng Tao, Xuelong Li, and Jie Yang. Patch alignment for dimensionality reduction. *Knowledge and Data Engineering, IEEE Transactions on*, 21(9):1299–1313, 2009.

[ZWT11] Wei Zhang, Xiaogang Wang, and Xiaoou Tang. Coupled information-theoretic encoding for face photo-sketch recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 513–520. IEEE, 2011.